

Exponential inequalities for the distributions of canonical U - and V -statistics of dependent observations

October 9, 2009

I. S. Borisov and N. V. Volodko

Abstract

The exponential inequalities are obtained for the distribution tails of *canonical* (degenerate) U - and V -statistics of an arbitrary order based on samples from a stationary sequence of observations satisfying φ -mixing.

Key words: *stationary sequence of random variables, φ -mixing, multiple orthogonal series, canonical U - and V -statistics.*

1. INTRODUCTION. PRELIMINARY RESULTS

The paper deals with estimates for the distribution tails of U - and V -statistics with canonical bounded kernels, based on samples of stationary observations under φ -mixing. The exponential inequalities obtained are a natural generalization of the classical Höfding inequality for the distribution tail of a sum of independent identically distributed bounded random variables. The approach of the present paper is based on the kernel representation of the statistics under consideration as a multiple orthogonal series (for detail, see [5, 9]). It allows to reduce the problem to more traditional estimates for the distribution tail of a sum of weakly dependent random variables.

Introduce basic definitions and notions.

Let X_1, X_2, \dots be a stationary sequence of random variables taking values in an arbitrary measurable space $\{\mathfrak{X}, \mathcal{A}\}$, with the common distribution F . In addition to the stationary sequence introduced above, we need an auxiliary sequence $\{X_i^*\}$ consisting of independent copies of X_1 .

Denote by $L_2(\mathfrak{X}^m, F^m)$ the space of measurable functions $f(t_1, \dots, t_m)$ defined on the corresponding Cartesian power of the space $\{\mathfrak{X}, \mathcal{A}\}$ with the corresponding product-measure and satisfying the condition

$$\mathbb{E}f^2(X_1^*, \dots, X_m^*) < \infty.$$

Definition 1. A function $f(t_1, \dots, t_m) \in L_2(\mathfrak{X}^m, F^m)$ is called *canonical* (or *degenerate*) if

$$\mathbb{E}_{X_k^*} f(X_1^*, \dots, X_m^*) = 0 \text{ a. s.} \quad (1)$$

for every k , where $\mathbb{E}_{X_k^*}$ is the conditional expectation given the random variables $\{X_i^*; i \leq m, i \neq k\}$.

Define a Von Mises statistic (or *V-statistic*) by the formula

$$V_n \equiv V_n(f) := n^{-m/2} \sum_{1 \leq j_1, \dots, j_m \leq n} f(X_{j_1}, \dots, X_{j_m}). \quad (2)$$

In the sequel, we consider only the statistics where the function $f(t_1, \dots, t_m)$ (the so-called *kernel* of the statistic) is canonical. In this case, the corresponding Von Mises statistic is also called *canonical*. For independent $\{X_i\}$, such statistics are studied during last sixty years (see the reference and examples of such statistics in [9]). In addition to *V-statistics*, the so-called *U-statistics* were studied as well:

$$U_n \equiv U_n(f) := n^{-m/2} \sum_{1 \leq i_1 \neq \dots \neq i_m \leq n} f(X_{i_1}, \dots, X_{i_m}). \quad (3)$$

Notice also that any *U-statistic* is represented as a finite linear combination of canonical *U-statistics* of orders from 1 to m . This representation is called *Höfding's decomposition* (see [9]).

For independent observations $\{X_i\}$, we give below a brief review of results directly connected with the subject of the present paper. In this connection, we would like to mention the results in [3, Theorem 1], [2, Proposition 2.2], [1, Theorem 7, Corollary 3], and [7, Theorem 3.3].

One of the first papers where exponential inequalities for the distribution tails of *U-statistics* are obtained, is the article by W. Höfding [8] although he considered nondegenerated *U-statistics* only. In this case, the value $(n - m)!/n!$ equivalent to n^{-m} as $n \rightarrow \infty$, is used as the normalizing factor instead of $n^{-m/2}$. In [8], the following statement is proved:

$$\mathbb{P}(U - \mathbb{E}U \geq t) \leq e^{-2kt^2/(b-a)^2}, \quad (4)$$

where

$$U = (n - m)!/n! \sum_{1 \leq i_1 \neq \dots \neq i_m \leq n} f(X_{i_1}, \dots, X_{i_m}),$$

$a \leq f(t_1, \dots, t_m) \leq b$ and $k = [n/m]$. In the case $m = 1$, inequality (4) is usually called *Höfdding's inequality* for sums of independent identically distributed bounded random variables. Notice that, in this case, the sums mentioned may be simultaneously considered as canonical or nondegenerate U -statistics.

In [3], an improvement of (4) was obtained for the case when there exists a splitting majorant of the canonical kernel under consideration:

$$|f(t_1, \dots, t_m)| \leq \prod_{i \leq m} g(t_i), \quad (5)$$

and the function $g(t)$ satisfies the condition

$$\mathbb{E}g(X_1)^k \leq \sigma^2 L^{k-2} k! / 2$$

for all $k \geq 2$. In this case, the following analogue of Bernstein's inequality holds:

$$\mathbb{P}(|V_n| \geq t) \leq c_1 \exp \left(-\frac{c_2 t^{2/m}}{\sigma^2 + L t^{1/m} n^{-1/2}} \right), \quad (6)$$

where the constants c_1 and c_2 depend only on m . Moreover, as noted in [3], inequality (6) cannot be improved in a sense.

It is clear that if $\sup_{t_i} |f(t_1, \dots, t_m)| = B < \infty$ then, in (6), one can set $\sigma = L = B^{1/m}$. Then it suffices to consider only the deviation zone $|t| \leq B n^{m/2}$ in (6) (otherwise the left-hand side of (6) turns into zero). Therefore, for all $t \geq 0$, inequality (6) yields the upper bound

$$\mathbb{P}(|V_n| \geq t) \leq c_1 \exp \left(-\frac{c_2}{2} (t/B)^{2/m} \right) \quad (7)$$

which is an analogue of Höfdding's inequality (4).

In [2], an inequality close to (6) is proved without condition (5), and relation (7) is given as a consequence. In [7], some refinement of (7) is obtained for $m = 2$, and in [1], the later result was extended to canonical U -statistics of an arbitrary order. The goal of the present paper is to extend inequality (7) to the case of stationary random variables under φ -mixing. For dependent observations, we do not yet know how to get more precise inequalities close to Bernstein's inequality (6), for unbounded kernels under some moment restrictions only.

2. MAIN RESULTS FOR WEAKLY DEPENDENT OBSERVATIONS

In the sequel, we assume that \mathfrak{X} is a separable metric space. Then the Hilbert space $L_2(\mathfrak{X}, F)$ has a countable orthonormal basis $\{e_i(t)\}$. Put $e_0(t) \equiv 1$. Using the Gram–Schmidt orthogonalization, one can construct an orthonormal basis in $L_2(\mathfrak{X}, F)$ containing the constant function $e_0(t) \equiv 1$. Then $\mathbb{E}e_i(X_1) = 0$ for every $i \geq 1$ due to orthogonality of all the other basis elements to the function $e_0(t)$. The normalizing condition means that $\mathbb{E}e_i^2(X_1) = 1$ for all $i \geq 1$.

In the sequel, we assume that the basis consists of uniformly bounded functions:

$$\sup_{i,t} |e_i(t)| \leq C. \quad (8)$$

It is well-known that the collection of the functions

$$\{e_{i_1}(t_1)e_{i_2}(t_2) \cdots e_{i_m}(t_m); \quad i_2, \dots, i_m = 0, 1, \dots\}$$

is an orthonormal basis of the Hilbert space $L_2(\mathfrak{X}^m, F^m)$. The kernel $f(t_1, \dots, t_m)$ can be decomposed by the basis $\{e_{i_1}(t_1) \cdots e_{i_m}(t_m)\}$ and represented as a series:

$$f(t_1, \dots, t_m) = \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} e_{i_1}(t_1) \cdots e_{i_m}(t_m), \quad (9)$$

where the series on the right-hand side converges in the norm of $L_2(\mathfrak{X}^m, F^m)$. Moreover, if the coefficients $\{f_{i_1, \dots, i_m}\}$ are absolutely summable then, due to the B. Levi theorem and the simple estimate $\mathbf{E}|e_{i_1}(X_1^*) \cdots e_{i_m}(X_m^*)| \leq 1$, the series in (9) converges almost surely with respect to the distribution F^m of the vector (X_1^*, \dots, X_m^*) . It is worth noting that, even in this case, we cannot extend this statement to the distribution of the vector with dependent coordinates (X_1, \dots, X_m) . Note also that $e_0(t)$ is absent in representation (9) because the kernel is canonical (for detail, see [5]).

Thus, by (9), we can represent the corresponding Von Mises statistic (2) as follows:

$$\begin{aligned} V_n &= n^{-m/2} \sum_{1 \leq j_1, \dots, j_m \leq n} f(X_{j_1}^*, \dots, X_{j_m}^*) \\ &= n^{-m/2} \sum_{1 \leq j_1, \dots, j_m \leq n} \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} e_{i_1}(X_{j_1}^*) \cdots e_{i_m}(X_{j_m}^*) \\ &= \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} n^{-1/2} \sum_{j=1}^n e_{i_1}(X_j^*) \cdots n^{-1/2} \sum_{j=1}^n e_{i_m}(X_j^*) \\ &= \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} S_n(i_1) \cdots S_n(i_m), \end{aligned}$$

where $S_n(i_k) := n^{-1/2} \sum_{j=1}^n e_{i_k}(X_j^*)$, $k = 1, \dots, m$.

In the present paper, we consider only stationary sequences $\{X_j\}$ satisfying φ -mixing condition. Recall the definition of this type of dependence. For $j \leq k$, denote by \mathfrak{M}_j^k the σ -field of all events generated by the random variables X_j, \dots, X_k .

Definition 2. A sequence X_1, X_2, \dots satisfies φ -mixing (or uniformly strong mixing) if

$$\varphi(i) := \sup_{k \geq 1} \sup_{A \in \mathfrak{M}_1^k, B \in \mathfrak{M}_{k+i}^\infty, \mathbb{P}(A) > 0} \frac{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|}{\mathbb{P}(A)} \rightarrow 0, \quad i \rightarrow \infty.$$

Remark 1. If $\{X_j\}$ satisfies φ -mixing with the coefficient $\varphi(k)$ then the sequence $f(X_1), f(X_2), \dots$ for every measurable function f also satisfies φ -mixing with the coefficient no more than $\varphi(k)$.

Introduce some additional restrictions on the mixing coefficient and the kernels of the statistics under consideration:

1. **(A)** $\sum_{i_1, \dots, i_m=1}^\infty |f_{i_1, \dots, i_m}| < \infty$ and $\varphi(k) \leq c_0 e^{-c_1 k^2}$, where $c_1 > 0$.
2. **(B)** There exists $\varepsilon > 0$ such that $\sum_{i_1, \dots, i_m=1}^\infty |f_{i_1, \dots, i_m}|^{1-\varepsilon} = c < \infty$; and

$$\varphi := \sum \varphi(k) < \infty.$$

The main results of the present paper are contained in the following two theorems.

Theorem 1. Let a canonical kernel $f(t_1, \dots, t_m)$ be continuous (in every argument) everywhere on \mathfrak{X}^m and let condition (8) be fulfilled. Moreover, if one of conditions (A) or (B) is fulfilled then the following inequality holds:

$$\mathbb{P}(|V_n| > x) \leq C_1 \exp \left\{ -C_2 x^{2/m} / B(f) \right\}, \quad (10)$$

where $C_2 > 0$ depends only on $\varphi(\cdot)$; in case (A),

$$B(f) := \left(C^m \sum_{i_1, \dots, i_m=1}^\infty |f_{i_1, \dots, i_m}| \right)^{2/m},$$

and in case (B),

$$B(f) := C^2 \left(\sum_{i_1, \dots, i_m=1}^\infty |f_{i_1, \dots, i_m}|^{1-\varepsilon} \right)^{\frac{2}{m(1-\varepsilon)}};$$

here the constant C is defined in (8).

Remark 2. Under condition (A), we may set $C_1 = 1$. Under condition (B), the value C_1 depends on the constants m , ε , c , and C . The dependence on the values c and C can be removed by considering “large enough” values of x , namely, satisfying the inequality

$$x^{2/m} \geq \varepsilon^{-1} 8m(1 - \varepsilon) e C^2 \varphi c^{2/(m(1-\varepsilon))}.$$

The following theorem is an analogue of statement (10) for U -statistics.

Theorem 2. Let the sequence X_1, X_2, \dots satisfy the following condition:

(AC) For every collection of pairwise distinct subscripts j_1, \dots, j_m , the distribution of $(X_{j_1}, \dots, X_{j_m})$ is absolutely continuous with respect to the distribution of (X_1^*, \dots, X_m^*) .

Moreover, if the basis $\{e_i(t)\}_{i \geq 0}$ satisfies restriction (8) and one of the two conditions (A) or (B) is valid then

$$\mathbb{P}(|U_n| > x) \leq C_1 \exp \left\{ -C_2 x^{2/m} / B(f) \right\}, \quad (11)$$

where, under condition (A), the constants C_1 and C_2 are the same as in Theorem 1, and under condition (B), the constant C_1 depends on m , ε , c , and C , and the constant C_2 depends on φ and m ; the value $B(f)$ is defined in Theorem 1.

3. PROOF OF THE THEOREMS

Proof of Theorem 1. Without loss of generality, we assume that a separable metric space \mathfrak{X} coincides with the support of the distribution F . The last means that \mathfrak{X} does not contain open balls with F -measure zero. Since all the basis elements $e_k(t)$ in (9) are continuous and uniformly bounded in t and k , due to Lebesgue’s dominated convergence theorem, the series in (9) is continuous if the coefficients f_{i_1, \dots, i_m} are absolutely summable. It is not difficult to see that, in this case, the equality in (9) turns into the identity on the all variables t_1, \dots, t_m because the equality of two continuous functions on an everywhere dense set implies their coincidence everywhere. So, in this case, one can substitute *arbitrarily dependent* observations for variables t_1, \dots, t_m in identity (9). Therefore, for *all elementary events*, the above-mentioned representation holds:

$$V_n(f) = \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} S_n(i_1) \cdots S_n(i_m), \quad (12)$$

where, as above, $S_n(i_k) := n^{-1/2} \sum_{j=1}^n e_{i_k}(X_j)$, $k = 1, \dots, m$.

(I) We first prove the theorem under condition (A). Consider an arbitrary even moment of the Von Mises statistic using representation (12):

$$\mathbb{E}V_n^{2N} = \sum_{i_1, \dots, i_{2mN}=1}^{\infty} f_{i_1, \dots, i_m} \cdots f_{i_{2mN-m+1}, \dots, i_{2mN}} \mathbb{E}S_n(i_1) \cdots S_n(i_{2mN}). \quad (13)$$

Further, we have

$$\begin{aligned} \left| \mathbb{E}S_n(i_1) \cdots S_n(i_{2mN}) \right| &\leq n^{-mN} \sum_{j_1, \dots, j_{2mN} \leq n} \left| \mathbb{E}e_{i_1}(X_{j_1}) \cdots e_{i_{2mN}}(X_{j_{2mN}}) \right| \\ &= n^{-mN} \sum_{r=1}^{2mN} \sum_{k_1 < \dots < k_r \leq n} \sum_{s_j(i): i \leq r, j \leq 2mN} \left| \mathbb{E}\nu_{k_1} \cdots \nu_{k_r} \right|, \end{aligned} \quad (14)$$

where

$$\nu_{k_i} = e_{i_1}^{s_1(i)}(X_{k_i}) \cdots e_{i_{2mN}}^{s_{2mN}(i)}(X_{k_i}), \quad s_j(i) \geq 0, \quad \sum_{j=1}^{2mN} s_j(i) > 0,$$

and

$$\sum_{i=1}^r \sum_{j=1}^{2mN} s_j(i) = 2mN.$$

Notice that, for a fixed r , the number of all collections of such $s_j(i)$ that i runs from 1 to r and j runs from 1 to $2mN$, coincides with the number of all different arrangements of $2mN$ indistinguishable elements in r cells when every cell must contain at least one element. It is well known that this value equals

$$C_{2mN-1}^{r-1} := \frac{(2mN-1)!}{(2mN-r)!(r-1)!}.$$

Lemma 1. *If $\{X_i\}$ satisfies φ -mixing and restriction (A) is valid then, for every collection $\{i_1, \dots, i_{2mN}\}$, the following inequality holds:*

$$\left| \mathbb{E}S_n(i_1) \cdots S_n(i_{2mN}) \right| \leq (\tilde{c}C^2mN)^{mN},$$

where \tilde{c} depends only on the mixing coefficient $\varphi(\cdot)$.

Proof. We will estimate every summand of the external sum over r in (14) (taking the normalization n^{-mN} into account). The approach is quite analogous to that in the proof of the corresponding assertion in [4, Lemma 4].

If $r \leq mN$ then the number of the collections $k_1 < \dots < k_r \leq n$ equals C_n^r and does not exceed $n^r \leq n^{mN}$. Hence,

$$n^{-mN} \sum_{k_1 < \dots < k_r \leq n} \sum_{s_j(i)} |\mathbb{E} \nu_{k_1} \dots \nu_{k_r}| \leq C^{2mN} C_{2mN-1}^{r-1}.$$

Now, let $r > mN$. Fix an arbitrary collection of $s_j(i)$ and consider the inner subsum in (14):

$$\sum_{k_{v_1} < \dots < k_{v_2} \leq n} |\mathbb{E} \nu_{k_{v_1}} \dots \nu_{k_{v_2}}|. \quad (15)$$

Here $1 \leq v_1 < v_2 \leq r$ and $v := v_2 - v_1 + 1$ is the multiplicity of the corresponding subsum, and the blocks ν_{k_i} are defined as before. Denote

$$\sum_{i=v_1}^{v_2} \sum_{j=1}^{2mN} s_j(i) = K(v_1, v_2).$$

Notice that, for $v_1 \leq l < v_2$, we have

$$K(v_1, l) + K(l+1, v_2) = K(v_1, v_2).$$

In the sequel, we will call ν_{k_i} a *short block* if

$$\sum_{j=1}^{2mN} s_j(i) = 1,$$

i. e., $\nu_{k_i} = e_{i_j}(X_{k_i})$ for some $1 \leq j \leq 2mN$. Prove the following statement: If the number of short blocks in the summands of subsum (15) is no less than $d \in \{0, 1, \dots, v\}$ then the following estimate is valid:

$$\sum_{k_{v_1} < \dots < k_{v_2} \leq n} |\mathbb{E} \nu_{k_{v_1}} \dots \nu_{k_{v_2}}| \leq (c_2 d)^{d/2-2} n^{v-d/2} C^{K(v_1, v_2)}, \quad (16)$$

where c_2 depends only on the mixing coefficient. Prove it by induction on d for all v_1 and v_2 such that $v \leq r$. First, let $d = 1$, i. e., the moments in (15) contain at least one short block. Denote it by ν_{k_l} , where $k_{v_1} \leq k_l \leq k_{v_2}$. In addition, set

$$\|\xi\|_t = (\mathbb{E} |\xi|^t)^{1/t}.$$

Then the following estimate holds:

$$\begin{aligned}
& \sum_{k_{v_1} < \dots < k_{v_2} \leq n} \left| \mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{v_2}} \right| \\
& \leq 2 \sum_{k_{v_1} < \dots < k_{v_2} \leq n} \varphi^{1/2}(k_{l+1} - k_l) \left\| \nu_{k_{v_1}} \cdots \nu_{k_l} \right\|_2 \left\| \nu_{k_{l+1}} \cdots \nu_{k_{v_2}} \right\|_2 \\
& \quad + \sum_{k_{v_1} < \dots < k_l \leq n} \left| \mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_l} \right| \sum_{k_{l+1} < \dots < k_{v_2} \leq n} \mathbb{E} \left| \nu_{k_{l+1}} \cdots \nu_{k_{v_2}} \right| \\
& \leq 2n^{v-1} C^{K(v_1, v_2)} \sum_{i=1}^{\infty} \varphi^{1/2}(i) \\
& \quad + 2n^{v_2-l} C^{K(l+1, v_2)} \sum_{k_{v_1} < \dots < k_l \leq n} \varphi^{1/2}(k_l - k_{l-1}) \left\| \nu_{k_{v_1}} \cdots \nu_{k_{l-1}} \right\|_2 \left\| \nu_{k_l} \right\|_2 \\
& \leq 2n^{v-1} C^{K(v_1, v_2)} \sum_{i=1}^{\infty} \varphi^{1/2}(i) \\
& \quad + 2n^{v_2-l} C^{K(l+1, v_2)} n^{l-v_1} C^{K(v_1, l)} \sum_{i=1}^{\infty} \varphi^{1/2}(i) \\
& = 4 \sum_{i=1}^{\infty} \varphi^{1/2}(i) n^{v-1} C^{K(v_1, v_2)} \\
& \leq 4 \sum_{i=1}^{\infty} \varphi^{1/2}(i) n^{v-1/2} C^{K(v_1, v_2)}
\end{aligned}$$

what required to be proved. The induction base is proved.

Now, let the inequality

$$\sum_{k_{v_1} < \dots < k_{v_2} \leq n} \left| \mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{v_2}} \right| \leq (c_2 z)^{z/2-2} n^{v-z/2} C^{K(v_1, v_2)}$$

hold for all the minimal possible numbers $z < d$ of the short blocks and all the multiplicities v , and the moments in (15) contain no less than d short blocks. Denote these blocks by $\nu_{k_{j_1}}, \dots, \nu_{k_{j_d}}$. Consider $d-1$ pairs of neighbor blocks of the type $\nu_{k_{j_s}}, \nu_{k_{j_s+1}}$, $s = 1, \dots, d-1$. Denote the differences between the subscripts in these pairs by t_1, \dots, t_{d-1} respectively. Among the summands in (15), select $d-1$ classes (in general, intersecting). We have

$$\sum_{k_{v_1} < \dots < k_{v_2} \leq n} \left| \mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{v_2}} \right| \leq R_1 + \dots + R_{d-1},$$

where the subsum R_s is taken over the set of subscripts

$$I_s := \left\{ (k_{v_1}, \dots, k_{v_2}) : k_{v_1} < \dots < k_{v_2} \leq n, \quad t_s = \max t_i \right\}.$$

We estimate every subsum R_s as follows:

$$\begin{aligned}
R_s &\leq 2 \sum_{I_s} \varphi^{1/2}(k_{j_s+1} - k_{j_s}) \left\| \nu_{k_{v_1}} \cdots \nu_{k_{j_s}} \right\|_2 \left\| \nu_{k_{j_s+1}} \cdots \nu_{k_{v_2}} \right\|_2 \\
&\quad + \sum_{k_{v_1} < \cdots < k_{j_s}} \left| \mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{j_s}} \right| \sum_{k_{j_s+1} < \cdots < k_{v_2}} \left| \mathbb{E} \nu_{k_{j_s+1}} \cdots \nu_{k_{v_2}} \right|. \tag{17}
\end{aligned}$$

Consider the first sum on the right-hand side of (17). We have

$$\begin{aligned}
&2 \sum_{I_s} \varphi^{1/2}(k_{j_s+1} - k_{j_s}) \left\| \nu_{k_{v_1}} \cdots \nu_{k_{j_s}} \right\|_2 \left\| \nu_{k_{j_s+1}} \cdots \nu_{k_{v_2}} \right\|_2 \\
&\leq 2C^{K(v_1, v_2)} \sum_{I_s} \varphi^{1/2}(t_s) \leq 2C^{K(v_1, v_2)} n^{v-(d-1)} \sum_{t_i: t_i \leq t_s} \varphi^{1/2}(t_s) \\
&\leq 2C^{K(v_1, v_2)} n^{v-d+1} \sum_{k=1}^n \varphi^{1/2}(k) k^{d-2} \leq 2C^{K(v_1, v_2)} n^{v-d+1} \sum_{k=1}^{\infty} \varphi^{1/2}(k) k^{d-2} \\
&\leq 2C^{K(v_1, v_2)} n^{v-d+1} c_0 \int_0^{\infty} e^{-c_1 t^2/2} t^{d-2} dt \\
&\leq \frac{c_0}{c_1^{(d-1)/2}} C^{K(v_1, v_2)} n^{v-d+1} \Gamma\left(\frac{d-1}{2}\right) \leq \frac{c_0 c_3}{c_1^{(d-1)/2}} C^{K(v_1, v_2)} n^{v-d/2} d^{d/2-4}.
\end{aligned}$$

The last inequality holds due to the evident fact that there exists a constant c_3 such that $\Gamma(t) \leq c_3 t^{t-4}$.

Now, consider the product of sums on the right-hand side of (17). Let the summands in the first of these sums contain d_1 short blocks selected above. Correspondingly, in the summands of the second sum, there are $d - d_1$ selected short blocks. By construction, we have $1 \leq d_1 \leq d - 1$, so, for both the sums, we can apply the induction assumption. So we have

$$\begin{aligned}
&\sum_{k_{v_1} < \cdots < k_{j_s}} \left| \mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{j_s}} \right| \sum_{k_{j_s+1} < \cdots < k_{v_2}} \left| \mathbb{E} \nu_{k_{j_s+1}} \cdots \nu_{k_{v_2}} \right| \\
&\leq (c_2 d_1)^{d_1/2-2} C^{K(v_1, j_s)} n^{j_s - v_1 + 1 - d_1/2} \\
&\quad \times \left(c_2 (d - d_1) \right)^{(d-d_1)/2-2} C^{K(j_s+1, v_2)} n^{v_2 - j_s - (d-d_1)/2} \\
&\leq (c_2 d)^{d/2-4} C^{K(v_1, v_2)} n^{v-d/2}.
\end{aligned}$$

Combining the estimates for the two sums on the right-hand side of (17), we have

$$R_s \leq (c_2 d)^{d/2-3} C^{K(v_1, v_2)} n^{v-d/2}.$$

Summing all the values R_s , we conclude that

$$\begin{aligned}
\sum_{k_{v_1} < \cdots < k_{v_2}} \left| \mathbb{E} \nu_{k_{v_1}} \cdots \nu_{k_{v_2}} \right| &\leq (d-1) (c_2 d)^{d/2-3} C^{K(v_1, v_2)} n^{v-d/2} \\
&\leq (c_2 d)^{d/2-2} C^{K(v_1, v_2)} n^{v-d/2}
\end{aligned}$$

what required to be proved. The last step is true only if $c_2 \geq 1$ but we can assume this condition to be satisfied without loss of generality.

Now, note that if $r > mN$ then, in the summands in (14), there are no less than $2(r - mN)$ short blocks. Thus, setting in (16) $v_1 := 1$, $v_2 := r$, $d := 2(r - mN)$, and $v := r$, we get the following estimate:

$$\begin{aligned} n^{-mN} \sum_{k_1 < \dots < k_r \leq n} \sum_{s_j(i)} \left| \mathbb{E} \nu_{k_1} \cdots \nu_{k_r} \right| &\leq C_{2mN-1}^{r-1} n^{-mN} (c_2 d)^{d/2-2} n^{v-d/2} C^{K(v_1, v_2)} \\ &\leq 2^{2mN-1} (c_2 2mN)^{mN-2} C^{2mN} \leq (8c_2)^{mN-1} C^{2mN} (mN)^{mN-2}. \end{aligned}$$

Summing over all r from 1 to $2mN$ in (14), we conclude that

$$\left| \mathbb{E} S_n(i_1) \cdots S_n(i_{2mN}) \right| \leq (8c_2)^{mN-1} C^{2mN} (mN)^{mN}.$$

The lemma is proved. \square

By this lemma, we estimate the even moment of the Von Mises statistic in (13):

$$\mathbb{E} V_n^{2N} \leq \left(\sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| \right)^{2N} (\tilde{c} C^2 mN)^{mN}.$$

Apply Chebyshev inequality:

$$\mathbb{P}(|V_n| > x) \leq \frac{\mathbb{E} V_n^{2N}}{x^{2N}} \leq \frac{\left(\sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| C^m \right)^{2N} (\tilde{c} mN)^{mN}}{x^{2N}}.$$

Set $N = \varepsilon x^{2/m}$ for some $\varepsilon > 0$. Then

$$\begin{aligned} \mathbb{P}(|V_n| > x) &\leq \frac{\left(\sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| C^m \right)^{2N} (\tilde{c} m \varepsilon)^{mN} x^{2N}}{x^{2N}} \\ &= \exp \left\{ \varepsilon m \ln(c_4 m \varepsilon) x^{2/m} \right\}, \end{aligned}$$

where

$$c_4 = \tilde{c} \left(\sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| C^m \right)^{2/m} = \tilde{c} B(f).$$

It is easy to verify that the multiplier $\varepsilon m \ln(c_4 m \varepsilon)$ reaches its minimum at the point $\varepsilon = (c_4 m e)^{-1}$, and this minimal value equals $-(c_4 e)^{-1}$. Then

$$\mathbb{P}(|V_n| > x) \leq \exp \left\{ -(\tilde{c} B(f))^{-1} x^{2/m} \right\}$$

what required to be proved.

(II) Now, prove the theorem under condition (B). Formulate the following auxiliary statement from [6, Proposition 5] adapted to our conditions:

Theorem 3. Let Y_1, Y_2, \dots be a stationary sequence of random variables taking values in \mathbb{R} and satisfying φ -mixing condition, and let $|Y_1| \leq C$ with probability 1. Then, for all $t > 0$, the following relation holds:

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_i - n\mathbb{E}Y_1\right| > t\right) \leq e^{1/e} \exp\left[\frac{-t^2}{16C^2 e \sum_{k=1}^{n-1} (n-k)\varphi(k)}\right]. \quad (18)$$

Set

$$\alpha_{i_1, \dots, i_m} = \begin{cases} |f_{i_1, \dots, i_m}|^{-\varepsilon}, & \text{if } f_{i_1, \dots, i_m} \neq 0, \\ \infty, & \text{if } f_{i_1, \dots, i_m} = 0. \end{cases}$$

We also agree that $0/0 = 0$ to make the following relation correct:

$$|f_{i_1, \dots, i_m} \alpha_{i_1, \dots, i_m}| = |f_{i_1, \dots, i_m}|^{1-\varepsilon}.$$

Obviously,

$$\sum_{i_1, \dots, i_m=1}^{\infty} \alpha_{i_1, \dots, i_m}^{-(1-\varepsilon)/\varepsilon} = \sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| \alpha_{i_1, \dots, i_m} = \sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m}^{1-\varepsilon} = c. \quad (19)$$

Further, evaluate the tail probability:

$$\begin{aligned} \mathbb{P}(|V_n(f)| > x) &= \mathbb{P}\left(\left|\sum_{i_1, \dots, i_m=1}^{\infty} f_{i_1, \dots, i_m} S_n(i_1) \cdots S_n(i_m)\right| > x\right) \\ &\leq \mathbb{P}\left(\sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m} S_n(i_1) \cdots S_n(i_m)| > c^{-1}x \sum_{i_1, \dots, i_m=1}^{\infty} |f_{i_1, \dots, i_m}| \alpha_{i_1, \dots, i_m}\right) \\ &\leq \sum_{i_1, \dots, i_m=1}^{\infty} \mathbb{P}\left(|S_n(i_1) \cdots S_n(i_m)| > c^{-1}x \alpha_{i_1, \dots, i_m}\right) \\ &\leq \sum_{i_1, \dots, i_m=1}^{\infty} \left(\mathbb{P}\left(|S_n(i_1)| > (c^{-1}x \alpha_{i_1, \dots, i_m})^{1/m}\right) + \cdots \right. \\ &\quad \left. + \mathbb{P}\left(|S_n(i_m)| > (c^{-1}x \alpha_{i_1, \dots, i_m})^{1/m}\right)\right). \end{aligned}$$

Taking into account condition (8) and the fact that, for every fixed basis function $e_i(t)$, the sequence $\{e_i(X_j)\}_j$ satisfies φ -mixing condition (Remark 1), we can set $Y_k = e_i(X_k)$ in Theorem 3 and use inequality (18) to estimate the probability $\mathbb{P}(|S_n(i)| > (c^{-1}x \alpha_{i_1, \dots, i_m})^{1/m})$. So, we have

$$\begin{aligned} \mathbb{P}(|S_n(i)| > (c^{-1}x \alpha_{i_1, \dots, i_m})^{1/m}) &\leq e^{1/e} \exp\left[\frac{-n(c^{-1}x \alpha_{i_1, \dots, i_m})^{2/m}}{16eC^2 \sum_{k=1}^{n-1} (n-k)\varphi(k)}\right] \\ &\leq e^{1/e} \exp\left[\frac{-n(c^{-1}x \alpha_{i_1, \dots, i_m})^{2/m}}{16eC^2 n \sum_{k=1}^{n-1} \varphi(k)}\right] \leq e^{1/e} \exp\left[\frac{-(c^{-1}x \alpha_{i_1, \dots, i_m})^{2/m}}{16eC^2 \varphi}\right]. \end{aligned}$$

If $\alpha_{i_1, \dots, i_m} = \infty$ then both the parts of this inequality are equal to zero. Note that the estimate above does not depend on the argument i of $S_n(i)$. Therefore, in the sequel, we will use the notation $S_n(1)$ only. Consider the multiple series from (19):

$$\sum_{i_1, \dots, i_m=1}^{\infty} \alpha_{i_1, \dots, i_m}^{-(1-\varepsilon)/\varepsilon} = c < \infty.$$

It follows from the series convergence that its elements can be order in decrease, and hence, the sequence $\{\alpha_{i_1, \dots, i_m}\}$ can be order in increase. Denote the well-ordered sequence by $\{\beta_i\}$. Obviously,

$$\sum_{i=1}^{\infty} \beta_i^{-(1-\varepsilon)/\varepsilon} = c,$$

and thus,

$$\beta_i \geq \left(\frac{i}{c}\right)^{\varepsilon/(1-\varepsilon)}.$$

So, taking the above-mentioned arguments into account, we can obtain the estimate

$$\begin{aligned} \mathbb{P}(|V_n(f)| > x) &\leq m \sum_{i=1}^{\infty} \mathbb{P}\left(|S_n(1)| > (c^{-1}x\beta_i)^{1/m}\right) \\ &\leq m \sum_{i=1}^{\infty} \mathbb{P}\left(|S_n(1)| > (c^{-1}x(ic^{-1})^{\varepsilon/(1-\varepsilon)})^{1/m}\right) \\ &\leq me^{1/e} \sum_{i=1}^{\infty} \exp\left[\frac{-(xi^{\varepsilon/(1-\varepsilon)}c^{-1/(1-\varepsilon)})^{2/m}}{16eC^2\varphi}\right] \\ &= me^{1/e} \sum_{i=1}^{\infty} \exp\left(-K(x)i^{2\varepsilon/(m(1-\varepsilon))}\right); \end{aligned}$$

here

$$K(x) = \frac{x^{2/m}}{16eC^2\varphi c^{2/(m(1-\varepsilon))}}.$$

Denote $\gamma = 2\varepsilon/(m(1-\varepsilon))$. We first consider such x that $\gamma K(x) \geq 1$ (it is clear that, in this case, $x \geq x_0$ for some x_0). We get

$$\sum_{i=1}^{\infty} \exp\left(-K(x)i^{\gamma}\right) \leq e^{-K(x)} + \int_1^{\infty} e^{-K(x)t^{\gamma}} dt = e^{-K(x)} + \frac{1}{\gamma} \int_1^{\infty} e^{-K(x)u} u^{1/\gamma-1} du.$$

If $\gamma \geq 1$ then

$$\frac{1}{\gamma} \int_1^{\infty} e^{-K(x)u} u^{1/\gamma-1} du \leq \frac{1}{\gamma} \int_1^{\infty} e^{-K(x)u} du = \frac{e^{-K(x)}}{\gamma K(x)} \leq e^{-K(x)}.$$

Otherwise, set $l = \lceil 1/\gamma - 1 \rceil + 1$. Notice that $l \leq K(x)$. Then

$$\begin{aligned} \frac{1}{\gamma} \int_1^\infty e^{-K(x)u} u^{1/\gamma-1} du &\leq \frac{1}{\gamma} \int_1^\infty e^{-K(x)u} u^l du \\ &= \frac{e^{-K(x)}}{\gamma K(x)} \left(1 + \frac{l}{K(x)} + \frac{l(l-1)}{K^2(x)} + \cdots + \frac{l!}{K^l(x)} \right) \\ &\leq \frac{(l+1)e^{-K(x)}}{\gamma K(x)} \leq (l+1)e^{-K(x)}. \end{aligned}$$

Thus, the theorem is proved for $x \geq x_0$. To do the statement of the theorem true for all nonnegative x , it suffices to increase C_1 so that the inequality

$$C_1 \exp \left\{ -C_2 x_0^{2/m} / B(f) \right\} \geq 1$$

holds. Then, for $x < x_0$, we obtain

$$\mathbb{P}(|V_n(f)| > x) \leq 1 \leq C_1 e^{-C_2 x_0^{2/m} / B(f)} \leq C_1 e^{-C_2 x^{2/m} / B(f)}.$$

Theorem 1 is proved. \square

Proof of Theorem 2. As noted above, the series in (9) converges almost surely with respect to the distribution of the vector (X_1^*, \dots, X_m^*) if the coefficients f_{i_1, \dots, i_m} are absolutely summable. It is clear that condition (AC) allows to claim the same on the distribution of the vector $(X_{j_1}, \dots, X_{j_m})$ for every *pairwise distinct* subscripts j_1, \dots, j_m .

Since, in the summation set in the definition of U -statistics, all the subscripts are pairwise distinct, we can substitute the series in (9) for the kernel in expression (3). So we obtain

$$U_n = \sum_{i_1, \dots, i_m=1}^\infty f_{i_1, \dots, i_m} \sum_{1 \leq j_1 \neq \dots \neq j_m \leq n} e_{i_1}(X_{j_1}) \cdots e_{i_m}(X_{j_m}).$$

Under condition (A), the proof repeats the previous one almost literally. Estimating the even moment of the statistic in the same way, we can obtain an upper bound for the multiple sum of mixed moments of the basis elements, which appear as a result of raising the U -statistic to the corresponding power. The difference between the right-hand side of (14) and this expression is in the absence of certain summands from (14) with multiple subscripts in the later multiple sum. Since all the summands on the right-hand side of (14) are non-negative, this is an upper bound for the above-mentioned mixed moment. Hence the moment

$$\left| \mathbb{E} \sum_{1 \leq j_1 \neq \dots \neq j_m \leq n} e_{i_1}(X_{j_1}) \cdots e_{i_m}(X_{j_m}) \cdots \sum_{1 \leq j_1 \neq \dots \neq j_m \leq n} e_{i_{2mN-m+1}}(X_{j_1}) \cdots e_{i_{2mN}}(X_{j_m}) \right|$$

does not exceed the expression obtained in Lemma 1. From here the statement of Theorem 2 follows in the case (A), with the same constants as those in Theorem 1.

Now, consider the situation when condition (B) is fulfilled. Notice that, by adding and subtracting the corresponding diagonal subspaces in the summation set, one can represent the expression

$$\sum_{1 \leq j_1 \neq \dots \neq j_m \leq n} e_{i_1}(X_{j_1}) \cdots e_{i_m}(X_{j_m}) \quad (20)$$

as a sum of products of the following values:

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n e_i(X_j), \quad \frac{1}{n} \sum_{j=1}^n e_{i_1}(X_j) e_{i_2}(X_j), \quad \dots, \quad \frac{1}{n^{k/2}} \sum_{j=1}^n e_{i_1}(X_j) \cdots e_{i_k}(X_j).$$

For convenience, denote the last of this sum as $S_n(i_1, \dots, i_k)$. Thus, we represent the U -statistic as a sum of Von Mises statistics (of the same or lesser orders and, perhaps, nondegenerate in case under consideration). Notice that the number of summands here depends only on m . It is worth noting that the first summand of this decomposition of the multiple sum in (20) is the product $S_n(i_1) \cdots S_n(i_m)$. Then we realize the proof in the same way as the proof of Theorem 1 for the case (II). The probability

$$\mathbb{P} \left(\left| \sum_{1 \leq j_1 \neq \dots \neq j_m \leq n} e_{i_1}(X_{j_1}) \cdots e_{i_m}(X_{j_m}) \right| > c^{-1} x \alpha_{i_1, \dots, i_m} \right)$$

is estimated by the sum of probabilities of the type

$$\mathbb{P} \left(\left| S_n(i_1^l, \dots, i_{k_1}^l) \cdots S_n(i_{m-k_s+1}^l, \dots, i_m^l) \right| > c^{-1} x \alpha_{i_1, \dots, i_m} / K(m) \right).$$

Here $K(m)$ is the above-mentioned number of summands which depends only on m ; $\{i_1^l, \dots, i_m^l\}$ is a permutation of the subscripts $\{i_1, \dots, i_m\}$, and $k_1 + \dots + k_s = m$. Estimate such a summand, for simplicity, omitting the superscripts and denoting by α the right-hand part of the inequality inside the last probability. Then

$$\begin{aligned} & \mathbb{P} \left(\left| S_n(i_1, \dots, i_{k_1}) \cdots S_n(i_{m-k_s+1}, \dots, i_m) \right| > \alpha \right) \\ & \leq \mathbb{P} \left(\left| S_n(i_1, \dots, i_{k_1}) \right| > \alpha^{k_1/m} \right) + \dots + \mathbb{P} \left(\left| S_n(i_{m-k_s+1}, \dots, i_m) \right| > \alpha^{k_s/m} \right). \end{aligned}$$

Using Theorem 3 and the fact that the value $S_n(i_1, \dots, i_k)$ (and consequently, its expectation) does not exceed $C^k/n^{k/2-1}$, we obtain the estimate

$$\begin{aligned} \mathbb{P}\left(\left|S_n(i_1, \dots, i_k)\right| > \alpha^{k/m}\right) &\leq e^{1/e} \exp \left[\frac{-\left(\alpha^{k/m} - C^k/n^{k/2-1}\right)^2 n^k}{16eC^{2k}n\varphi} \right] \\ &= e^{1/e} \exp \left[\frac{-\left(\alpha^{k/m} - C^k/n^{k/2-1}\right)^2 n^{k-1}}{16eC^2\varphi C^{2(k-1)}} \right]. \end{aligned}$$

It is easy to see that, for every fixed x , beginning from some n , the argument of the last exponent does not exceed the value

$$\frac{-\alpha^{2/m}}{16eC^2\varphi}.$$

Then it remains to repeat the proof of Theorem 1 for this case. Theorem 2 is proved. \square

References

- [1] R. Adamczak, “Moment inequalities for U -statistics,” *Ann. Probab.* **34** (6), 2288–2314 (2006).
- [2] M. A. Arcones and E. Giné, “Limit theorems for U -processes,” *Ann. Probab.* **21** (3), 1494–1542 (1993).
- [3] I. S. Borisov, “Approximation of distributions of von Mises statistics with multidimensional kernels,” *Sibirsk. Mat. Zh.* **32** (4), 20–35 (1991) [*Siberian Math. J.* **32** (4), 554–566 (1991)].
- [4] I. S. Borisov and A. A. Bystrov, “Limit theorems for the canonical von Mises statistics with dependent data,” *Sibirsk. Mat. Zh.* **47** (6), 1205–1217 (2006) [*Siberian Math. J.* **47** (6), 980–989 (2006)].
- [5] I. S. Borisov and N. V. Volodko, “Orthogonal series and limit theorems for canonical U - and V -statistics of stationary connected observations,” *Mat. Trudy* **11** (1), 25–48 (2008) [*Sib. Adv. Math.* **18** (4), 244–259 (2008)].
- [6] J. Dedecker and C. Prieur, “New dependence coefficients. Examples and applications to statistics,” *Probab. Theory Related Fields* **132** (2), 203–236 (2005).

- [7] E. Giné, R. Latała, and J. Zinn, “Exponential and moment inequalities for U -statistics,” in Proceedings of High Dimensional Probability II (Seattle, WA, 1999); Progr. Probab. **47**, 13–38 (2000) (Boston: Birkhäuser).
- [8] W. Höffding, “Probability inequalities for sums of bounded random variables,” J. Amer. Statist. Assoc. **58**, 13–30 (1963).
- [9] V. S. Korolyuk and Yu. V. Borovskikh, *Theory of U -Statistics* (Naukova Dumka, Kiev, 1989) [Kluwer Academic Publ., Dordrecht, 1994].
- [10] V. V. Petrov, *Sums of independent random variables* (Nauka, Moscow, 1972) [Springer-Verlag, New York, 1975].